

## **ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА**

на дисертаційну роботу Вавіленкової А. І.

«Теоретичні основи автоматичного аналізу і синтезу логіко-лінгвістичних моделей текстових документів»,

подану на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.06 – Інформаційні технології

**Актуальність теми.** Щорічний приріст інформації призводить до того, що людина не в змозі сприйняти її традиційними засобами. Для покращення автоматизованої обробки інформації потрібно вдосконалювати форму її подання. Складність і неосяжність різних суттєвих аспектів предметної області дослідження вимагає поєднання різнопланових моделей, технологій та інструментарію для ефективної програмної реалізації рішень з оптимізації пошуку інформації. Один із підходів до створення такої методології – це виявлення спільного інваріантного ядра та створення відповідного теоретичного логіко-понятійного базису. У свою чергу, основною проблемою інформаційного пошуку є підвищення ефективності знаходження релевантних документів для задоволення інформаційних потреб. Це можливо тільки шляхом застосування інформаційних технологій змістовного аналізу електронних документів, на що і спрямовані дослідження дисертаційної роботи Вавіленкової А. І. «Теоретичні основи автоматичного аналізу і синтезу логіко-лінгвістичних моделей текстових документів».

У сучасному світі знання поступово стають основною цінністю, одним із головних ресурсів людського суспільства. Тому актуальними на сьогодні є різноманітні технології вилучення знань.

Більшість пошукових систем використовують статистичні методи оцінки, ідентичні алгоритми пошуку за ключовими словами, доповнення запитів шаблонами, логічними операторами, спеціальними словами для пошуку в метаданих та ін. Ціннішими виявляються якісна підтримка морфології різних мов та врахування відстані між словами, проте вони не спрацьовують для великих розмірів даних. Це відбувається тому, що статистичні методи не в змозі впоратися з розпізнаванням омонімів і багатозначних слів через

неможливість розуміння сенсу запиту. Тому особлива увага зараз приділяється формалізованим підходам, що дають змогу створювати системи з високою інформаційною зв'язністю, тематичною однорідністю, достатньо високою структурованістю та якісним інформаційним наповненням. Тоді на заміну звичайному інформаційному пошуку приходять автоматизований експертний підбір найбільш адекватних інформаційних ресурсів із контекстним підбором матеріалів.

Розроблена Вавіленковою А. І. інформаційна технологія порівняльного аналізу електронних текстових документів поєднує у собі методи лінгвістичного аналізу з математичним апаратом логіки предикатів. Це представляє собою новий підхід до вирішення актуальної сьогодні проблеми змістовного аналізу інформації.

**Ступінь обґрунтованості, достовірність та новизна наукових положень, висновків та рекомендацій** забезпечується глибиною дослідження методів лінгвістичного аналізу та аналізу даних, коректним використанням математичного апарату і підтверджується проведеною оцінкою ефективності роботи створеної системи на експериментальних даних.

У дисертаційній роботі слід відзначити такі наукові результати:

1. Вперше створено математичний апарат для формального опису електронного текстового документу на базі логіки предикатів, що дає можливість структурувати текстову інформацію.

2. Вперше сформульовано умови тотожності логіко-лінгвістичних моделей речень природної мови та способи їх виявлення, розроблено принципи синтезу логіко-лінгвістичних моделей речень природної мови, що базуються на виявленні засобів змістовного зв'язку у текстових документах.

3. Вперше створено абстрактні моделі для формального опису логічних зв'язків між частинами текстових документів та їх геометричні інтерпретації, що стало підґрунтям для формування загальної форми логіко-лінгвістичної моделі електронного текстового документу.

4. Вперше розроблено методи автоматичного визначення логічної тотожності складних та простих предикатів, які входять до складу логіко-лінгвістичних моделей електронних текстових документів.

5. Вперше розроблено інформаційну технологію автоматичного порівняльного аналізу електронних текстових документів з використанням логіко-лінгвістичного моделювання.

**Практична цінність.** Усі теоретичні положення, які запропоновані в дисертаційній роботі, доведені до реально діючих комп'ютерних програм. На основі розробленої інформаційної технології створена система порівняльного аналізу електронних текстових документів, яка дозволяє підвищити ефективність пошуку збігів.

Результати дисертаційного дослідження впроваджено у науково-дослідну роботу Національного авіаційного університету кафедри комп'ютеризованих систем управління «Комп'ютерна технологія порівняльного аналізу електронних текстів» для створення проекту багатомовного пошукового серверу. Також результати дисертаційного дослідження використовуються у навчальному процесі Національного авіаційного університету Міністерства освіти і науки України. Отримані результати дисертаційного дослідження призначені для використання у галузі інформаційних технологій та комп'ютерної лінгвістики, що підтверджено актами впровадження.

**Повнота викладення результатів дисертаційної роботи.** Автор повністю відобразив основні положення дисертаційної роботи у 65 публікаціях, у тому числі 1 одноосібній монографії, 1 навчальному посібнику з грифом МОН України, 33 одноосібних наукових статтях, серед яких 20 статей у наукових фахових виданнях ДАК України, 8 у закордонних профільних журналах, 18 статей індексуються у міжнародних науково-метричних базах, 26 тезах доповідей на науково-технічних конференціях, 1 методичних рекомендаціях та 3 авторських свідоцтвах на комп'ютерні програми.

Дисертація оформлена згідно з вимогами ДАК України, а автореферат загалом відповідає змісту дисертації.

### **Зауваження щодо змісту дисертації.**

1. Виходячи із припущення, що подібність моделей типу (1) – (4), що розроблені автором у другому розділі, індукує подібність текстів, для яких вони побудовані, дисертант пропонує застосування таких моделей для задачі пошуку плагіату (розділ 4). Однак, питання типу відношення між логіко-лінгвістичною моделлю та відповідним їй текстом, а саме чи наявна еквівалентність, які аспекти текстової інформації втрачаються при побудові моделі, а які, навпаки, формалізуються, залишається відкритим і недослідженим автором.

2. Для порівняння моделей автором розроблено та описано алгоритм, який на кожному кроці виконує ряд модифікацій над моделлю, проте еквівалентність перетворень не доведена. Враховуючи зазначені зауваження, виникає питання адекватності розробленої моделі у сенсі представлення синтаксичних та семантичних аспектів тексту, для якого вона побудована.

3. У п'ятому розділі представлені результати порівняння розробленої автором системи перевірки тексту на плагіат із існуючими. Зважаючи на відсутність таких формальних складових методики порівняння систем, як: характеристики текстової вибірки; формального опису критеріїв оцінки якості; аргументації вибору класу текстових документів, критеріїв оцінки систем, результати порівняння не є показовими і не можуть бути застосовані для формулювання твердження про якість розробленої системи.

4. Необхідно було чітко окреслити різницю між запропонованою логіко-лінгвістичною моделлю речення природної мови та класичним поняттям предикату.

5. Варто зауважити, що у роботі зустрічається неузгодженість термінології, деякі синтаксичні та стилістичні помилки (стор. 18, 20, 21, 59, 60), наявність неаргументованих тверджень щодо особливостей згаданих методів (стор. 27, 49, "описані вище методи пошуку синонімів дієві з тією чи іншою ймовірністю лише для слів і не можуть бути застосовані до речень природної мови без застосування додаткових алгоритмів") та існуючих

текстових моделей (стор. 41, "жодна з існуючих текстових моделей не дає можливості відобразити змістовні зв'язки").

**Висновки.** Вище вказані зауваження не зменшують загальну позитивну оцінку роботи. Автор досягнув поставленої мети – підвищив якість процесу автоматичного порівняння електронних текстових документів шляхом використання вперше розроблених теоретичних і прикладних основ побудови інформаційних технологій.

Все це дає підстави стверджувати, що дисертаційна робота Вавіленкової Анастасії Ігорівни є завершеним науковим дослідженням, яке присвячене вирішенню актуальної науково-прикладної задачі. Дисертаційна робота повністю відповідає вимогам ДАК України щодо докторських дисертацій (згідно пп. 13 «Порядку присудження наукових ступенів і присвоєння вченого звання старшого наукового співробітника»), а Вавіленкова Анастасія Ігорівна заслуговує присудження їй наукового ступеня доктора технічних наук за спеціальністю 05.13.06 – «Інформаційні технології».

Офіційний опонент  
декан факультету інформатики  
Національного університету  
"Києво-Могилянська академія",  
доктор фізико-математичних наук, професор

М. М. Глибовець

